Adaptive coarse space construction and nonlinear smoothers for heterogeneous Stokes problems DI13C-2434

Jed Brown jedbrown@mcs.anl.gov, Mark Adams, Matt Knepley, Barry Smith

Multigrid in lithosphere and mantle dynamics

Heterogeneous Stokes problems appear in various forms throughout geodynamics, often coupled to viscoelasticity, viscoplasticity, and porous media flow. As a bottleneck of many high-resolution studies, robust and efficient Stokes solvers are needed. These methods are necessarily multilevel and require accurate coarse representations of operators. The problems arising in lithosphere dynamics are challenging for standard methods due to multiscale structures creating long-range interaction through thin structures that are difficult to accurately represent using conventional coarse spaces.

Nonlinearity: plasticity and phase change

Strong material nonlinearities such as plasticity cause methods based on global linearization, such as Newton and Picard, to require many iterations. Nonlinear multigrid avoids global linearization, leading to faster convergence rates when effective nonlinear smoothers are available. With a nonlinear smoother, we naturally want to build interpolation and the coarse operator without global assembly of a fine-grid operator. Unfortunately, traditional geometric multigrid does not accurately interpolate low-frequency modes and rediscretized coarse operators are notoriously inaccurate in highly heterogeneous cases. A subdomain-centric coarse grid construction only involves solving local problems, thus allowing it to be updated only in regions with rapidly-changing nonlinearities.

Matrix-free for performance

Assembled sparse matrices have long been a preferred representation for PDE operators, but are a remarkably poor fit for modern hardware due to memory bandwidth requirements. A matrix-vector product computed using an assembled matrix cannot have an arithmetic intensity higher than $1/4$, leaving modern floating point hardware severely under-utilized.

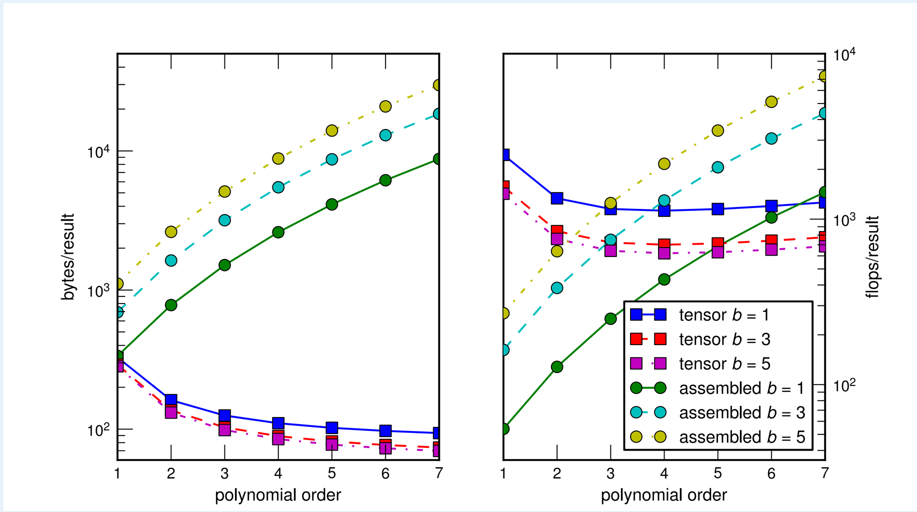


Figure: Relative cost in memory bandwidth and flops to apply linearized PDE operator arising in p-version finite element discretization of nonlinear PDEs with $b = 1, 3, 5$ degrees of freedom per node.

Processor	BW (GB/s)	Peak (GF/s)	Balanced AI (F/B)
Sandy Bridge 6-core	21*	150	7.2
Magny Cours 16-core	42*	281	6.7
Blue Gene/Q node	43	205	4.8
Tesla M2050	144	515	3.6
Kepler K20	250	1310	5.2

Table: Balanced arithmetic intensity (flops/byte) for several architectures.

The τ formulation for multiscale modeling

The Full Approximation Scheme is a naturally nonlinear multigrid algorithm that allows flexible incorporation of multilevel information.

- classical formulation: "coarse grid *accelerates* fine grid solution"
- τ formulation: "fine grid improves accuracy of coarse grid"
- To solve $Nu = f$, recursively apply

$$\text{pre-smooth} \quad \tilde u^h \leftarrow S^h_{\text{pre}}(u^h_0, f^h)$$

$$\text{solve coarse problem for } u^H \quad N^H u^H = \underbrace{I^H_h f^h}_{f^H} + \underbrace{N^H \hat I^H_h \tilde u^h - I^H_h N^h \tilde u^h}_{\tau^H_h}$$

$$\text{correction and post-smooth} \quad u^h \leftarrow S^h_{\text{post}}\left(\tilde u^h + I^h_H (u^H - \hat I^H_h \tilde u^h), f^h\right)$$

I^H_h	residual restriction	$\hat I^H_h$	solution restriction
I^h_H	solution interpolation	$f^H = I^H_h f^h$	restricted forcing
$\{S^h_{\text{pre}}, S^h_{\text{post}}\}$	smoothing operations on the fine grid		

- At convergence, $u^{H*} = \hat I^H_h u^{h*}$ solves the τ-corrected coarse grid equation $N^H u^H = f^H + \tau^H_h$, thus τ^H_h is the "fine grid feedback" that makes the coarse grid equation accurate.
- τ^H_h is *local* and need only be recomputed where it becomes stale.

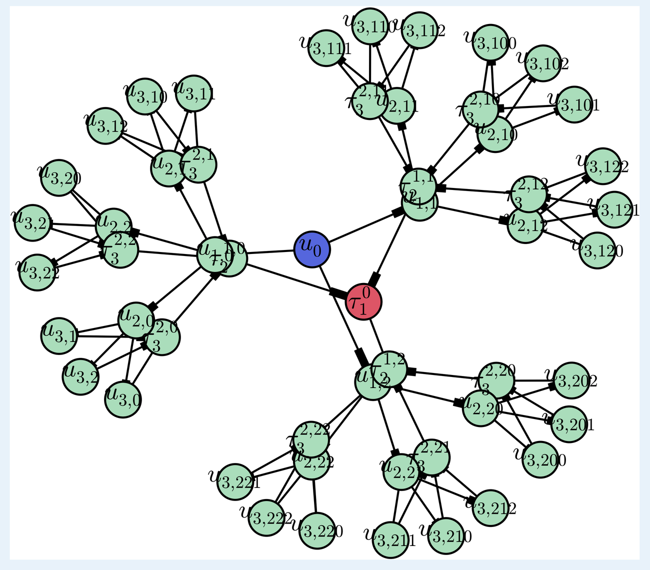
Subdomain-centric matrix-free coarsening

Objective: construct robust interpolation and coarse grid operator using only (a) local neighbor information, (b) application of local nonlinear operator, (c) point-block diagonal of principle linearization, and (d) application of triangular distribution operator or splitting [3] for saddle points.

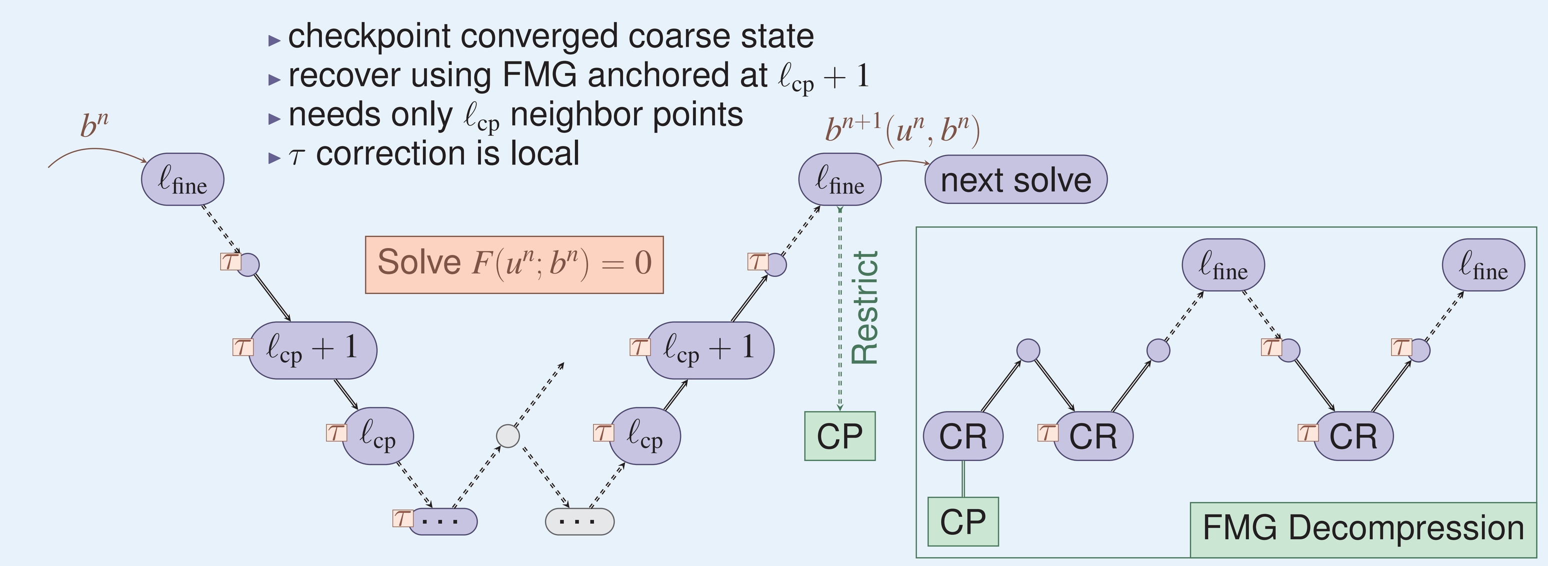
1. Select subdomains to become "coarse elements", add minimal stable node set to preliminary set of coarse dofs C.
2. If available, add approximate null space to set of "low-energy" modes B that must be approximated accurately.
3. Use compatible relaxation with point-block preconditioned polynomial smoother to determine deficiencies of initial coarse basis.
4. Enrich C by adding poorly-converging points.
5. Optimize energy of local basis functions by computing partition of coarse space B on the boundary, then (approximately) harmonically extending to subdomain interior.
6. Optionally, use (non-local) bootstrap cycle [1] to improve B.

Low-communication cycling

The τ formulation removes communication from all levels except the coarsest. Instead of starting and ending on the fine grid, a cycle starts and ends on the coarse grid. The figure shows the dependency graph of 3-level multigrid cycle that computes the correction τ^0_1 (red) on the coarse grid equation starting with coarse grid state u_0 (blue). A traditional multigrid cycle which has "horizontal" dependencies at every level.

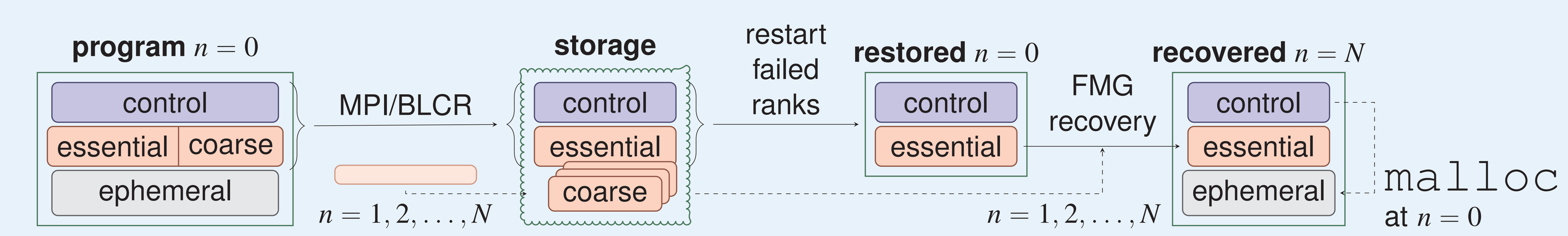


Multiscale compression and decompression



- Fine state u^{h*} recovered *locally* from converged coarse state $u^{H*} = \hat I^H_h u^{h*}$
- Normal multigrid cycles visit all levels moving from $n \to n+1$
- FMG recovery only accesses levels finer than ℓ_{CP}
- Only neighborhood of desired region needed during decompression
- Lightweight checkpointing for transient adjoint computation
- Postprocessing applications, e.g., in-situ visualization at high temporal resolution in part of the domain

Local decompression and resilience



control contains program stack, solver configuration, etc.

essential program state that cannot be easily reconstructed: time-dependent solution, current optimization/bifurcation iterate

ephemeral easily recovered structures: assembled matrices, preconditioners, residuals, Runge-Kutta stage solutions

- Essential state at time/optimization step n is inherently globally coupled to step $n-1$ (otherwise we could use an explicit method)
- *Coarse* level checkpoints are orders of magnitude smaller, but allow rapid recovery of essential state
- FMG recovery needs only nearest neighbors

Status

Proof-of-concept compatible relaxation and subdomain coarsening implemented using PETSc, similar robustness to modern smoothed aggregation. Low-communication implementation and use of more efficient data structures for local decomposition in progress. Merging subdomain-centric approach with PCGAMG (algebraic multigrid infrastructure), along with accessible user hooks for customization.

References

- A. Brandt, J. Brannick, K. Kahl, and I. Livshits. Bootstrap AMG. *SIAM Journal on Scientific Computing*, 33(2):612–632, 2011.
- A. Brandt and B. Diskin. Multigrid solvers on decomposed domains. *Contemporary Mathematics*, 157:135–135, 1994.
- J. Brown, M. G. Knepley, D. A. May, L. C. McInnes, and B. F. Smith. Composable linear solvers for multiphysics. In *Proceedings of the 11th International Symposium on Parallel and Distributed Computing (ISPDC 2012)*, pages 55–62. IEEE Computer Society, 2012. Also available as Argonne National Laboratory preprint ANL/MCS-P2017-0112.

This work was supported by the Department of Energy Office of Advanced Scientific Computing Research.